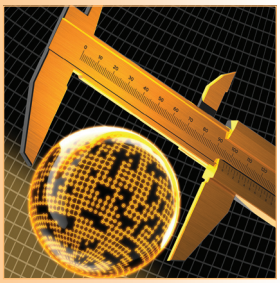
2016-08

Empirical Study of Router IPv6 Interface Address Distributions

Rohrer, Justin



Empirical Study of Router IPv6 Interface Address Distributions

IPv6 is an important component of the Internet's continued growth and evolution. It has grown exponentially and now carries nontrivial amounts of production traffic. Less well-understood is IPv6's topology and the way in which providers are using their IPv6 address allocations. Rather than relying on passive measurements or heuristics, the authors use uniform active probing; executing ICMP-Paris traceroute probes to an address in each /48 in all /32's advertised in the global IPv6 routing table (approximately 400 million traces). At this granularity, they characterize the distribution of IPv6 interface addresses in the wild, and find significant differences among providers and regions.

Justin P. Rohrer, Blake LaFever, and Robert Beverly
US Naval Postgraduate School

While the rate and prevalence of IPv6[1] adoption varies among different macro measurement methods,[2] today IPv6 carries a nontrivial fraction of production Internet traffic. Significant prior work has examined the adoption and evolution of IPv6, such as client IPv6 support,[3] prevalence of IPv6 Domain Name System (DNS) records,[2] dual-stack infrastructure,[4] and IPv4 versus IPv6 path performance.[5] Less attention, however, has been placed on IPv6 topology inference and network mapping.

Efforts to understand the Internet's IPv6 topology have largely focused on autonomous system (AS) connectivity via Border Gateway Protocol (BGP) control plane analysis[5] — an important, but coarse-grained view. By contrast, router-level topologies can reveal critical structure and operation within an AS. However, discovery and analysis of router and router interface topologies via active probing have thus far been limited, due in part to the fundamental difficulty posed by the sheer size of IPv6's 128-bit address space. Although Internet-wide active topology probing (such as traceroute) is performed regularly across the IPv4 address space,[6,7] it isn't clear how to conduct similar active topology scans for the entire IPv6 Internet. A current state-of-the-art, active IPv6 topology collection platform, the Center for Applied Internet Data Analysis's Archipelago (CAIDA's Ark),[8] conducts continuous Internet Control Message Protocol (ICMP)-Paris traceroute[9] probes to routed IPv6 destinations from a distributed set of vantage

Published by the IEEE Computer Society
1089-7801/16/$33.00 © 2016 IEEE

points. To make active probing feasible, CAIDA presently probes two addresses within each globally advertised IPv6 BGP prefix in each round of probing: the ::1 address, and a random address. While this is an intuitive strategy to balance probing cost and time with expected coverage, to the best of our knowledge, neither its completeness nor its soundness have been rigorously examined.

Keeping this in mind, here we seek to inform two closely-related questions regarding Internet IPv6 topology: first, how are IPv6 providers using and subnetting their address allocations; and second, how can active measurement platforms more effectively and efficiently sample IPv6 topology? The basis of our analysis is a uniform ICMP-Paris traceroute probing of each /48 within all /32 prefixes advertised in the global BGP table (thus, exhaustive probing at a /48 granularity; 216 probes per /32). This first-of-its-kind dataset of approximately 400 million traces from 26 vantage points provides a valuable approximation of ground-truth to understand current IPv6 subnetting and allocation practice in the wild. While more granular (longer mask) IPv6 subnetting exists, /48's were the recommended allocations to customers from a provider's allocation[10] for a decade before it became obsolete.[11] Thus, while we haven't exhaustively probed all possible subnets, we believe /48's represent a reasonable compromise between completeness and probing volume/time. Our contributions thus include the following:

- an analysis of active Paris-traceroute probing of an address in each /48 within all globally advertised /32 IPv6 prefixes (this dataset, gathered in collaboration with CAIDA, is publicly available[12]);
- a characterization of Internet-wide IPv6 allocations, subnetting, and adherence to recommended best common operational practices;
- an analysis of per-provider and per-regional differences in IPv6 subnetting; and
- the distribution of discovered IPv6 interfaces.

Our hope is that this work serves to inform both the development of future efficiency-optimized active IPv6 probing algorithms, as well as the community's understanding of provider use of IPv6 address allocations.

Background

The growth, use, and adoption of IPv6 has been extensively measured and studied. Recently, Jakub Czyz and his colleagues found significant differences in the adoption of IPv6 in a longitudinal study across 10 different datasets, each representing a different facet of IPv6 (for example, use of IPv6 in the DNS, routing, and traffic).[2] Rather than presenting a broad study of adoption, we seek to more deeply understand a single aspect — IPv6 subnetting — in a single, Internet-wide snapshot.

Prior work examines IPv6 topology, but is largely limited to the AS-level topology as observed in BGP routing announcements. For example, Amogh Dhamdhere and his colleagues examine the congruence of IPv4 and IPv6 AS paths, IPv6 AS path lengths, and the most central IPv6 ASes.[5] In contrast, our work studies properties of the interface-level IPv6 address allocation, and statistical properties revealed via active probing.

Related to our work is the spatial classification of IPv6 addresses observed by a large content distribution network (CDN).[13] By clustering the addresses of Web clients that access the CDN, inference can be made on the ways in which providers are allocating addresses and subnets to clients. Our work is largely complementary to this prior study: rather than opportunistically relying on passive traffic (such as clients that access the CDN) our active probing helps eliminate possible sample bias. However, our technique limits us to understanding the addressing at a /48 granularity, while passive techniques can reveal more fine-grained details.

To the best of our knowledge, CAIDA performs the only continuously maintained active IPv6 topology discovery platform built on its Ark platform.[8] Each vantage point in Ark takes routed IPv6 prefixes (as viewed from the global BGP table) as input, and probes the following: the ::1 station address within each prefix; and a random station address within each prefix. Probes consist of ICMP-Paris traceroutes[9] performed by the scamper packet prober.[14]

We're motivated in part by previous work that performed active topology probing of the IPv6 Internet by proposing heuristics and techniques to cope with the address space's size.[15,16] A central problem, however, is obtaining a basis for evaluating the performance of such intelligent IPv6 active probing — without ground-truth of the possible topology that could be discovered, only relative metrics are possible when evaluating topology probing systems. As such, we hope our work serves to inform the development of future efficient IPv6 active network mapping algorithms.

In addition, our work sheds some light onto the way in which providers are using IPv6 addresses for infrastructure, and insight into the subnets they might be allocating. Such operational insight is interesting with respect to published best common operating practice guidelines,[11,17-19] which recommend the following, for instance: dedicating the first or last /48 per region to number infrastructure, numbering point-to-point interfaces out of /64 prefixes, not subnetting on non-nibble boundaries, and creating subnet prefixes of equal size. We find distinct evidence in practice of both adherences to, and deviations from, these recommendations.

Methodology

Because we take a provider-centric view, our work centers on studying prefixes at the /32 granularity (that is, IPv6 prefixes with 32 bits of network mask). We start with the set of globally advertised IPv6 BGP prefixes, as visible from routeviews (www.routeviews.org), and limit our examination to advertised /32 prefixes. Future work should subdivide larger prefixes (those with masks less than 32) into constituent /32s.

To perform the active topology probing, we rely on scamper,[14] an advanced packet prober that implements a variety of traceroute methods. We use scamper to send Time-to-Live (TTL) limited ICMP6 probes, where the probe headers are formed using the Paris traceroute technique[9] such that each probe takes the same path, even over flow-balanced paths. The probes elicit ICMP6 TTL-exceeded messages, where the source address corresponds to the router's interface used to reach the prober.[20] Thus, we recover the set of router interface IPv6 addresses on the forward path toward the destination.

To interrogate every /48 in the set of globally advertised /32 IPv6 prefixes visible from routeviews, we use 26 IPv6-capable vantage points from the CAIDA Ark[6] infrastructure. The Ark vantage points are physically globally distributed, as well as connected to a diverse set of IPv6 networks. We distribute the task to issue IPv6 ICMP-Paris traceroutes to the `::1` address in every /48 (one traceroute to each destination) across these vantage points. In total, our data includes one-time probing toward approximately 408 Million /48s. We chose the `::1` address due to its general popularity as the subnet gateway router interface address. While the tools and platform are the same, this methodology is quite different from CAIDA's routine IPv6 probing, discussed in the "Background" section, which uses scamper-equipped Ark nodes to probe one random address in each of the roughly 16,000 advertised prefixes (regardless of prefix length) every three days.[8]

Scamper terminates probing toward a destination after reaching a gap limit of five successive unresponsive hops, or if a loop is detected (where the same IPv6 address is observed in response to two different probe TTLs). These settings mirror CAIDA's default configuration for their continuous probing, thereby helping to mitigate comparison bias between the two datasets.

As we mentioned, the entire dataset collected and analyzed in this article is publicly available from CAIDA.[12] So not only should our results be repeatable, we hope this dataset enables future research and analysis of IPv6 topology.

Probing Details

Due to the volume of probing required, probing was performed from 13 November 2014 to 1 March 2015. The set of prefixes to probe is based on the information available at the beginning of this time period, at which time there were 6,162 /32s. The assignment of probes to vantage-points is pseudo-random. The list of 32-bit prefixes is shuffled randomly and then split into batches of 110 prefixes. The 110 × 216 /48s in each batch then is shuffled randomly and split into fixed-size chunks. Each chunk is assigned round-robin to a particular vantage point for probing.

We distributed the probing work among vantage points to spread the probing load and decrease the total time required. While different vantage points will take different paths to reach a particular destination, each vantage point probes a random and large subset of the total destination set. For a given traceroute, newly discovered interfaces primarily come from hops past the vantage point's local neighborhood, which is already well-discovered over the course of probing. Thus, the net impact of using multiple vantage points on the discovered interface graph is likely to be negligible.

Out of the 407,793,780 probes issued, only 137,235 (0.03 percent) reached their destination — an unsurprising result due to the currently sparse IPv6 Internet population. We also determined that 35,118,396 probes (8.6 percent) were terminated due to loop detection, and another

Table 1. Topology discovery statistics – uniform probing versus existing Center for Applied Internet Data Analysis (CAIDA) methodology.

Feature	Uniform	CAIDA (All)	CAIDA (/32's)
Traces	407,793,780	1,059,058	302,531
– Destination reached	137,235 (0.03%)	119,052 (11.5%)	33,728 (11.1%)
– Looped	35,118,396 (8.6%)	81,779 (7.7%)	19,331 (6.4%)
– Gapped	203,837,347 (49.7%)	527,822 (49.8%)	145,687 (48.2%)
Interfaces	128,804	57,455	29,882
Edges	267,647	144,311	68,801

203,837,347 (50 percent) due to gap limiting (too many unresponsive hops). The remainder of the probes terminated with some other ICMP6 unreachable code. All of this probing results in a set of 240,155 unique IPv6 addresses, belonging to 152,481 unique /48 subnets, in 4,763 unique /32s prefixes, belonging to 4,173 different ASes. These 240,000 IPv6 addresses consist of 137,235 unique responsive destinations, and 128,804 unique intermediate router IPv6 interface addresses (where a responsive destination in one trace might also appear as an intermediate hop in a second trace).

Address Population

This population study is primarily concerned with the allocation and use of IPv6 addresses for numbering router interfaces. IPv6 router address discovery is important to building topology maps and understanding critical infrastructure. While it's difficult to discover responsive IPv6 hosts (our traceroutes empirically yield only a 0.03 percent response rate), by sending probes to destinations uniformly across each /32, we hope to elicit responses from as many routers along the forwarding paths as possible. For this reason, we also consider the intermediate traceroute hops in our results. This reveals additional interfaces both within the /32 being probed, as well as in other /32s in use by transit providers. Naturally, we discover some of the same interfaces appearing in many traceroutes, but duplicates are removed and only unique interface addresses retained in the population. However, the effect of using all responding interfaces in our results is that the effective sampling rate isn't uniform across all /32s; those in use by transit providers receive a much higher sampling rate. To account for this, in the following section we consider not only aggregated statistics, but individual /32 distributions, and their contribution to the overall statistical results.

Results

Here, we present a relative comparison of our probing against CAIDA's current IPv6-wide active topology probing, analyze the distribution of interfaces within /32s, and characterize difference among /32s observed in the wild.

Relative Comparison

We first seek to understand the relative difference between our uniform probing, and the current state-of-the-art collection from CAIDA.[8] Our objective is to obtain a sense of how much topology information is gathered today, versus how much could be obtained through more exhaustive and/or intelligent probing.

Table 1 compares our uniform probing against one cycle of CAIDA's production IPv6 probing collected on 1–2 March 2015 from 26 vantage points (both our method and CAIDA's production method use the same vantage points). In addition, we analyze a subset of the CAIDA traces that corresponds to exactly those /32's we probe. Because CAIDA probes destinations within all routed prefixes, this restricted dataset is a more meaningful comparison by using the same set of /32 prefixes.

We observe approximately the same fraction of gap limited and looped traces across all three datasets, although more traces loop in our uniform probing, presumably because we target destinations for which there's no more specific route. The CAIDA traces reach a much higher fraction of destinations due to the fact that there's frequently a "::1" address associated with each routed prefix, as opposed to our probing of largely unused space.

We then examine the number of unique interfaces (router interface IP addresses) and edges (IP address pairs in successive traceroute hops) obtained in each dataset. CAIDA's probing of the /32s finds only 23 percent of the number of interfaces our uniform probing discovers,

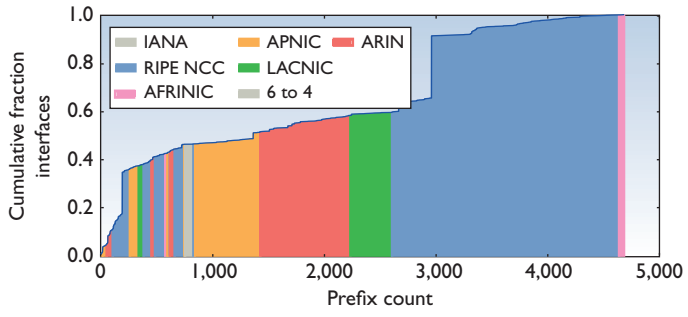


Figure 1. All of the IPv6 interfaces discovered, sorted by the /32 prefix. Vertical jumps in the plot show concentrations of interfaces discovered in particular /32 prefixes. Réseaux IP Européens Network Coordination Center (RIPE NCC) is dominant, both in the number of active prefixes and in the fraction of interfaces discovered. (AFRINIC = African Network Information Center, APNIC = Asia-Pacific Network Information Center, ARIN = American Registry for Internet Numbers; IANA = Internet Assigned Numbers Authority, and LACNIC = Latin America and Caribbean Network Information Center.)

and only 26 percent of the number of edges from uniform probing. However, this coverage comes at the cost of greater than 1,000 times more probes.

These high-level findings suggest that there's a significant amount of currently undiscovered (on a per-probing round basis) topology, yet the cost of discovering this topology with current methodology is prohibitively high.

Distribution of Interfaces within /32s

We next examine the distribution of sources (router interfaces) replying to our probing (inclusive of intermediate router hops), as organized by the /32 to which the response's IPv6 source belongs. Looking at the number of unique unicast IPv6 router addresses within each /32, we see a wide variance (see Figure 1). Here, the *y*-axis is the cumulative fraction of total discovered unique IPv6 interfaces, while the *x*-axis is simply a sequential identifier for each /32 (assigned in increasing order of network prefix). Note that the total number of observed /32s is fewer than 6,162 (the number of /32s we probe), because our probes didn't elicit responses from every /32 targeted. Further, we receive responses from /32s not in the original set of 6,162. Vertical jumps in the plot show concentrations of interfaces discovered in particular /32 prefixes. We also show via colored banding the regional Internet registry (RIR) to which each prefix belongs. This lets us see the dominance of Réseaux IP Européens (RIPE), both in the number of active prefixes and in the fraction of interfaces discovered. Table 2 lists the specific /32 prefixes in which the most interfaces were discovered. Comparing Table 2 to Figure 1, we can easily see the contributions of the densely populated /32s from Bitcanal and XS4ALL, as well as some of the other large interface blocks from major providers. Surprisingly, the majority of the ASes represented in Table 2 aren't tier-1 transit providers, and some have relatively low IPv6 AS rank.[21] (We based this information on the CAIDA IPv6 Org Rank dataset from 1 September 2014, which is the most recent available at the time of this writing.)

In Figure 2 we show the same data, this time with the *x*-axis aligned by the /32 prefix itself (as opposed to the sequential index of the /32). This view lets us observe the relative population of the few /8s from which allocations are currently distributed, with `2000::/8` and `2a00::/8` making up the lion's share, and `2600::`, `2400::/8`, and `2800::/8` a distant third, fourth, and fifth, respectively. In terms of regional allocations, RIPE, American Registry for Internet Numbers (ARIN), Asia-Pacific Network Information Center (APNIC), Latin America and Caribbean Network Information Center (LACNIC), and African Network Information Center (AFRINIC) all contribute to the `2000::/8` block, while the sole `2400::/8` allocation belongs to APNIC, all three `2600::` allocations belong to ARIN, and the only allocation from `2800::/8` is to LACNIC. The large number of interfaces in the `2a00::/8` block are in RIPE allocations, with a few AFRINIC interfaces next door in the `2c00::/8` allocation.

For the last component of our look at the overall population of /32s, we again rearrange the *x*-axis, this time sorting the prefixes from high to low by their interface count (see Figure 3). This plainly shows 90 percent of the observed interfaces coming from 10 percent of the prefixes.

Example /32 Interface Distributions

Having surveyed all of the /32s corresponding to each discovered interface, we then drill down to examine the distribution of responding interfaces within individual /32s. Figure 4 shows 10 of the top 12 /32 prefixes by census, and it's clear that IPv6 address allocation schemes vary

Table 2. Top /32 prefixed by number of unique interface addresses discovered.

Prefix	Interfaces	Autonomous system (AS) no.	Organization	Organizational rank
2a00:4c87::	62,392	197426	Bitcanal	7,281
2001:980::	40,860	3265	XS4ALL	2,217
2a00:f42::	7,869	43447	Orange	116
2001:470::	6,886	6939	Hurricane Electric	9
2406:e000::	5,873	23655	Snap Internet	525
2001:288::	4,503	1659	TANet	1,739
2001:4dd0::	2,708	8422	NetCologne	552
2001:6f8::	2,656	4589	Easynet	397
2001:668::	2,356	3257	Tinet	4
2001:550::	2,183	174	Cogent	2
2001:1900::	2,172	3356	Level 3	1
2001:7f8::	2,065	6695	DE-CIX	7,281

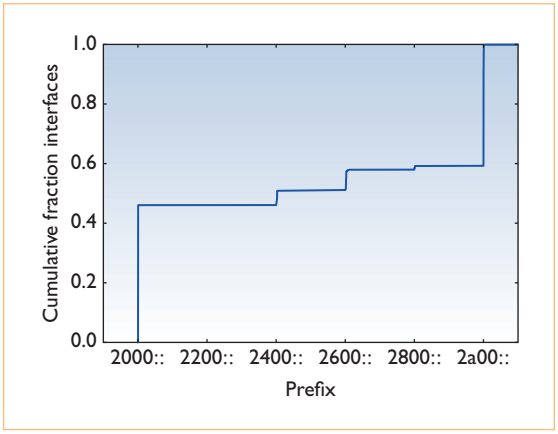


Figure 2. All of the IPv6 interfaces discovered, indexed by the /32 prefix. This view lets us observe the relative population of the few /8s from which allocations are currently distributed.

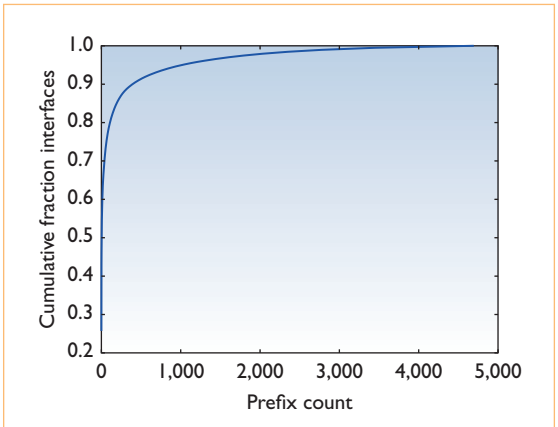


Figure 3. All of the IPv6 interfaces, with a reverse-sorted number of interfaces per /32 (sorting from high to low). Ninety percent of the observed interfaces come from 10 percent of the prefixes.

widely. Within AS43447 (ORANGE-PL) we find nearly 8,000 interfaces uniformly distributed between 2a00:f42:0:: and 2a00:f42:2000::, or one per /48, while the other 47,000 /48s are unpopulated. Hurricane electric on the other hand shows thousands of interfaces within the first and second /48 (2001:470:0:: and 2001:470:1::), and another grouping of thousands of interfaces between 2001:470:1f04:: and 2001:470:1f14::. Hurricane also has many interfaces distributed across the remainder of this /32. Looking further down the graph, we see a number of /32s (belonging to Tinet, Cogent, DE-CIX, and Level 3), where more than 1,000 interfaces appear in the first /48 and far fewer or no interfaces are discovered in the rest of the /32. We observe that this distribution seems to reflect current best practices for IPv6 address allocation, which prescribes allocating subnets from the beginning of the allocation.[17] Other /32s (belonging to Easynet, TANet, and Snap Internet) have blocks of responding interfaces more widely distributed across their /32. Last, we see a /32 belonging to NetCologne, with more than 2,000 interfaces responding in the 2001:4dd0:ff00:: /48 at the upper end of the /32, and with negligible allocations elsewhere in the range. This set of /32s illustrates

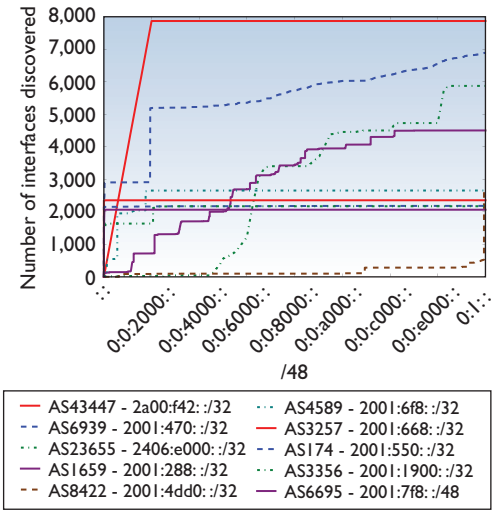


Figure 4. Example distribution of addresses from several /32 prefixes. Ten of the top 12 /32 prefixes are shown, and the IPv6 address allocation schemes vary widely.

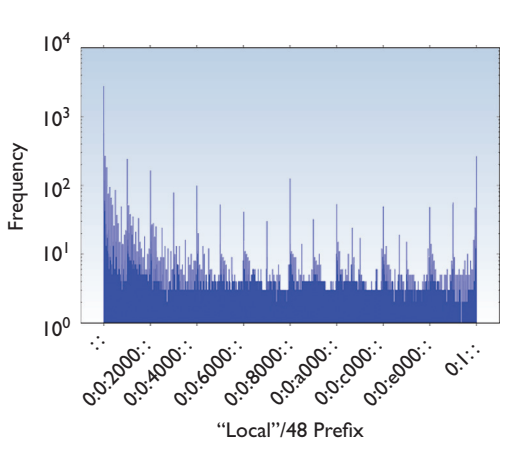


Figure 5. Frequency of occurrence of all /48 local prefix subnets.

the wide variety of IPv6 address allocation strategies observed in the wild.

We don't plot the top two /32 prefixes (listed in Table 2), because nearly every /48 is represented, resulting in interface counts that are an order of magnitude higher than the rest of the top 10 and appear simply as linear diagonal lines when plotted. It's not clear why these two ASes have so many interfaces responding.

While this sampling of the top responding /32s highlights the wide variance in addressing strategies, what we're really interested in discovering

Table 3. Most population /48 local subnets.

/48	Subnets
X:X:0::	2,756
X:X:1::	1,038
X:X:2::	571
X:X:3::	323
X:X:100::	267
X:X:ffff::	263
X:X:4::	247
X:X:10::	246
X:X:1000::	240
X:X:200::	182

Table 4. Most common 33–48 prefix bits in addresses.

/48	Interfaces
X:X:0::	25,572
X:X:1::	5,401
X:X:2::	3,603
X:X:ff00::	3,160
X:X:200::	2,759
X:X:100::	2,432
X:X:ffff::	1,651
X:X:900::	1,467
X:X:4::	1,279
X:X:5::	1,161

is if there are commonalities between different organizations' address-allocation schemes. For this we look to several plots showing local prefix distributions. Not to be confused with the IPv6 prefix reserved for link-local addressing, the local prefixes we refer to here are the 33rd through 48th bits of the network address. These are local subnet prefixes given an assumed /32 prefix allocation. We start with a histogram of /48 local prefix bit patterns shown in Figure 5, along with Table 3, which shows the exact values for the top 10 peaks from the histogram. We see that more than half of the /32s probed respond from the X:X:0::/48, and nearly one-fourth respond from the X:X:1::/48, which is consistent with the sample /32s we observe in Figure 4. Perhaps more interesting is the distinct spike we observe at X:X:8000::/48 and the somewhat lower spike at each /36 increment (X:X:1000::/48, X:X:2000::/48, X:X:3000::/48, and so on). We

think this is likely indicative of hierarchical subnetting (/34 and /36) taking place within the /32s, but could also reflect human preference for memorable subnet prefixes. We defer distinguishing the root cause of these patterns to future work.

We further break down the distribution of interfaces by RIR assignment in Figure 6, which shows significant differences between the regions. Table 5 gives a detailed breakdown of all interfaces discovered by RIR allocation. In comparison with Figure 4, we observe some similar patterns, in that the majority of the RIRs carry a large fraction of their interfaces within the first /48. RIPE is the most notable exception to this, but we find that the RIPE distribution is dominated by two /32s (`2a00:4c87::/32` and `2001:980::/32`) that contribute 103,252 of RIPE's 171,308 interfaces. In general, this reinforces the notion that the relative sparsity of the IPv6 Internet infrastructure, combined with the scale of individual allocations, allow a few actors to dominate the statistics of the whole. Additional comparisons between Table 5 and Table 2 show that it's likely a few major ASes dominating the /48 distribution in each of the regions.

Note that for Figures 5 and 6, we count only unique /48 bit patterns, not individual interfaces. The reason for this is that the interface count in the `X:X:0::/48` is so large (Table 4) that it made distinguishing patterns in the lower volume /48s difficult, even with a log-scale plot. Given that our probing strategy was only comprehensive to the granularity of /48 prefixes, it makes sense to aggregate to that level for statistical analysis. That being said, we also want to understand why there's such a disproportionately large number of interfaces recorded with the `X:X:0::/48` bit pattern (keep in mind that we only probed one address in each /48; discovering 10 interfaces for every responding `X:X:0::/48` was unexpected), so we perform further analysis of the paris-traceroutes where they were discovered. We find that on average every trace with at least one such interface contains five interfaces matching this pattern, and they appear in the middle of the traces. Looking at the /32 prefixes from the interfaces in question, we find that the vast majority belong to major transit ASes (Tinet, Cogent, Orange, Hurricane Electric, and Nippon Telegraph and Telephone [NTT]) and some large ISPs (Time Warner and Comcast). Comparing these observations with Figure 4 leads us to the conclusion

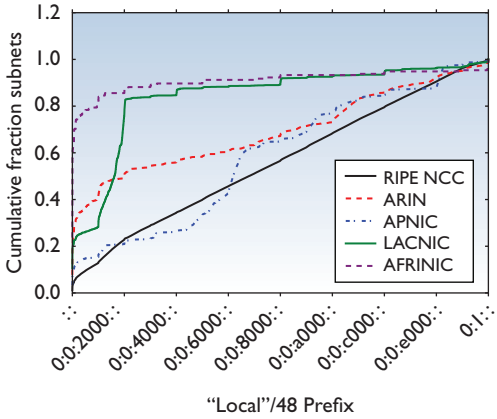


Figure 6. The /48 local prefix distribution by regional Internet registry (RIR). There are significant differences between the regions.

that large ISPs are more likely to assign IPv6 addresses densely from the start of their /32 allocation. A more comprehensive examination of the allocation of individual router interface host addresses is available elsewhere.[22]

Observations and Analysis

At its heart, our study is a population survey; however, it yields interesting insights and presents numerous avenues for follow-on study.

One unexpected result of this study is that the top 12 most populous /32s we discovered aren't primarily composed of the largest IPv6 customer-cone AS rank networks. Several are represented in our list, but they're offset by an equal number of low-ranked organizations. Looking at these most populous /32s reveals widely differing allocation schemes. Some have interfaces spread quasi-uniformly throughout the range (apparently by preference, not necessity), while others have them clustered exclusively at the beginning or end of the range. However, when we look only at the distribution of /48s, aggregated across all /32s, some highly distinct patterns become visible. Interfaces are clustered at common subnet boundaries — for example, /33 (`X:X:8000::`), /34 (`X:X:4000::`, `X:X:c000::`), /35 (`X:X:2000::`, `X:X:6000::`, and so on), and /36 (`X:X:1000::`, `X:X:3000::`, `X:X:5000::`, and so on). At the same time, it's also clear that human preference plays a large role, with 1, 2, 3, 100, ffff, 4, and 10 outranking any of the aforementioned prefixes in popularity (also observed in Matthew Gray's work[22]). We also note that our traceroute-based

Table 5. Interface distribution by RIR.

RIR	Unique /32s	Unique /48s	Interfaces
RIPE	2,448	130,246	171,308
ARIN	934	7,142	38,119
APNIC	734	12,336	25,770
LACNIC	417	2,473	4,412
AFRINIC	63	194	449
6to4	75	79	80
IANA	3	3	4

methodology primarily collects addresses assigned to outward-facing router interfaces, which might skew the results we observe.

Although our methodology results in a reasonable sample size, when compared with the potential population of a single /32, it's still quite small. This allows overall statistics to be dramatically skewed by one or two ASes and requires us to examine the results at a finer granularity. Specifically, Bitcanal and XS4ALL both have responding interface counts an order of magnitude higher than any of the tier-1 service providers, and account for over 40 percent of our total results. This dominates any aggregate statistic they're included in, particularly those for the RIPE RIR, to whose allocation they both belong. This is definitely an area we would like to investigate further, both to find out why those particular ASes respond in this way, as well as to determine aggregate analysis metrics that are meaningful in the context of the massive IPv6 population disparity that currently exists, and is likely to continue to exist, for many years to come.

By its nature and design, this work has raised as many new questions and research directions as it has answered. For instance, our exhaustive probing provides only a snapshot in time. It would be valuable to perform a longitudinal study to understand how the infrastructure IPv6 address population distributions shift over time. Further, while validation is difficult, we plan to solicit feedback on our findings from willing providers where possible.

Due to the volume of probing required, we also chose to limit this study to the granularity of /48s, but it's well-known that major providers are allocating subnets at a finer granularity than this, so more granular probing might also be beneficial, whether wide-scale or on selected prefixes. At the other end of the scale, the present work doesn't consider prefixes larger than /32 (for example, with masks greater than 32). While there are many more prefixes with 32 bits of mask than those with masks less than 32 bits, these larger prefixes represent large providers that we haven't yet fully characterized. We do this so as to scope our probing effort, while facilitating comparisons among provider /32s. In the future, we plan to subdivide these large aggregates into their constituent /32s, and include them in our uniform probing.

While our ultimate goal is understanding the IPv6 topology, this work is largely restricted to characterizing the population of IPv6 addresses discovered via exhaustive active probing. Future work can examine the same dataset[12] to make stronger topological inferences. For instance, if two probes to contiguous /48s are topologically congruent, this could imply that the /48s are part of a larger routed aggregate. In contrast, if paths to these two /48s reveal different router interfaces, we could more strongly infer that there exist distinct routed subnets for these two destinations.

In previous work, we showed the benefits of intelligent probing algorithms to improve efficiency (number of interfaces discovered/number of traceroutes required) in the IPv4 Internet,[23] but have struggled to apply these same algorithms to IPv6. We hope that the results of this exhaustive probing study can lead to more efficient IPv6-optimized probing algorithms. Last, we again take note of the large impact on our population statistics caused by only two /32s (belonging to Bitcanal in Portugal and XS4ALL in the Netherlands) that appeared to have nearly every /48 populated. Neither of these organizations is a highly ranked ISP (by the IPv6 customer-cone metric), so it strikes us as odd that they would have an order-of-magnitude larger interface population responding than organizations like Level 3 and Hurricane Electric. We would like to explore this finding further, to identify the allocation techniques in use at those organizations.

Acknowledgments

We thank the anonymous reviewers for their valuable comments, Young Hyun for his assistance in executing the exhaustive probing, Kimberly C. Claffy ("kc claffy") for her support, and CAIDA for the Ark infrastructure. This work was supported in part by US National Science Foundation grant CNS-1111445 and US Department of Homeland Security Cyber Security Division contract N66001-2250-58231. Views and conclusions are those of the authors and

shouldn't be interpreted as representing the official policies, either expressed or implied, of the US government.

References

1. S. Deering and R. Hinden, *Internet Protocol, Version 6 (IPv6) Specification*, IETF RFC 2460, Dec. 1998; www.ietf.org/rfc/rfc2460.txt.
2. J. Czyz et al., "Measuring IPv6 Adoption," *Proc. ACM Sigcomm Conf.*, 2014, pp. 87–98.
3. S. Zander et al., "Mitigating Sampling Error when Measuring Internet Client IPv6 Capabilities," *Proc. ACM Internet Measurement Conf.*, 2012, pp. 87–100.
4. R. Beverly and A. Berger, "Server Siblings: Identifying Shared IPv4/IPv6 Infrastructure via Active Fingerprinting," *Proc. 16th Conf. Passive and Active Network Measurement*, 2015, pp. 149–161.
5. A. Dhamdhere et al., "Measuring the Deployment of IPv6: Topology, Routing and Performance," *Proc. ACM Internet Measurement Conf.*, 2012, pp. 537–550.
6. Center for Applied Internet Data Analysis (CAIDA), *Archipelago (Ark) Measurement Infrastructure*, Mar. 2015; www.caida.org/projects/ark.
7. H.V. Madhyastha et al., "iPlane: An Information Plane for Distributed Services," *Proc. 7th Symp. Operating Systems Design and Implementation*, 2006, pp. 367–380.
8. CAIDA, *The CAIDA UCSD IPv6 Topology Dataset*, 2015; www.caida.org/data/active/ipv6_allpref_topology_dataset.xml.
9. B. Augustin et al., "Avoiding Traceroute Anomalies with Paris Traceroute," *Proc. ACM Internet Measurement Conf.*, 2006, pp. 153–158.
10. Internet Architecture Board and Internet Eng. Steering Group (IAB/IESG), *IAB/IESG Recommendations on IPv6 Address Allocations to Sites*, IETF RFC 3177, Sept. 2001; www.ietf.org/rfc/rfc3177.txt.
11. T. Narten, G. Huston, and L. Roberts, *IPv6 Address Assignment to End Sites*, IETF RFC 6177, Mar. 2011; www.ietf.org/rfc/rfc6177.txt.
12. Y. Hyun, *CAIDA IPv6 Routed /48 Topology Dataset*, 2015; www.caida.org/data/active/ipv6_routed_48_topology_dataset.xml.
13. D. Plonka and A. Berger, "Temporal and Spatial Classification of Active IPv6 Addresses," *Proc. ACM Sigcomm Internet Measurement Conf.*, 2015, pp. 509–522.
14. M. Luckie, "Scamper: A Scalable and Extensible Packet Prober for Active Measurement of the Internet," *Proc. ACM Internet Measurement Conf.*, 2010, pp. 239–245.
15. R. Barnes, R. Altmann, and D. Kerr, "Mapping the Great Void: Smarter Scanning for IPv6," *Internet Statistics and Metrics Analysis Workshop*, presentation, 2012; www.caida.org/workshops/isma/1202/slides/aims1202_rbarnes.pdf.
16. F. Gont and T. Chown, *Network Reconnaissance in IPv6 Networks*, IETF RFC 7707, Mar. 2016; www.ietf.org/rfc/rfc7707.txt.
17. C. Grundemann, A. Hughes, and O. DeLong, *Best Current Operational Practices – IPv6 Subnetting (v1)*, Global Network Eng. Comm., Feb. 2012; http://nabcop.org/index.php?title=IPv6_Subnetting.
18. B. Carpenter and S. Jiang, *Emerging Service Provider Scenarios for IPv6 Deployment*, IETF RFC 6036, Oct. 2010; www.ietf.org/rfc/rfc6036.txt.
19. K. Chittimaneni et al., *Enterprise IPv6 Deployment Guidelines*, IETF RFC 7381, Oct. 2014; www.ietf.org/rfc/rfc7381.txt.
20. A. Conta, S. Deering, and M. Gupta, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*, IETF RFC 4443 (updated by RFC 4884), Mar. 2006; www.ietf.org/rfc/rfc4443.txt.
21. M. Luckie et al., "AS Relationships, Customer Cones, and Validation," *Proc. ACM Sigcomm Internet Measurement Conf.*, Oct 2013, pp. 243–256.
22. M. Gray, "Discovery of IPv6 Router Interface Addresses via Heuristic Methods," master's thesis, Naval Postgraduate School, Sep. 2015; http://calhoun.nps.edu/handle/10945/47265.
23. G. Baltra, R. Beverly, and G.G. Xie, "Ingress Point Spreading: A New Primitive for Adaptive Active Network Mapping," *Proc. 15th Conf. Passive and Active Network Measurement*, 2014, pp. 56–66.

Justin P. Rohrer is a research assistant professor in the Department of Computer Science at the US Naval Postgraduate School (NPS), where he leads the Tactical Networked Communication Architecture Design (TanCAD) lab. His research focuses on tactical networking, including resilient and survivable mobile transport and routing protocols, and measuring network resilience. Rohrer has a PhD (with honors) in electrical engineering from the University of Kansas. Contact him at jprohrer@nps.edu.

Blake LaFever is a lieutenant commander in the United States Navy. His research interests include military cyber defense, information assurance, and network operations. LaFever has an MS in cyber systems and operations from the US Naval Postgraduate School. Contact him at bwlafeve@cmand.org.

Robert Beverly is an assistant professor in the Department of Computer Science at the US Naval Postgraduate School. His research in computer networks focuses on network architecture, large-scale traffic analysis and data mining, measurement, and security. Beverly has a PhD in electrical engineering and computer science from the Massachusetts Institute of Technology. Contact him at rbeverly@nps.edu.